

1. Project Title

Major Depression: Stage 1 Genomewide Association in Population-Based Samples

2. Abstract

Eight methodological issues have led to the historical difficulties in dissecting the molecular genetics of MDD (a common, complex trait of enormous public health significance). GAIN inherently corrects two problems – inadequate statistical power and sparse, candidate-gene based genotyping.

Additionally, first, most MDD samples are inhomogeneous in subject ancestry and are concatenated from disparate studies. Second, most MDD collections are highly selected samples of convenience containing unknown but potentially disastrous biases. Third, “controls” used in most association studies are highly dissimilar to cases in ascertainment and assessment *and* may not be at low likelihood for MDD. Fourth, minimal phenotypes are usually available – typically self-reported and exclude biological markers that can identify more homogeneous subgroups. Fifth, replication has historically been conducted external to a study and not within the study. Most critically, sixth, MDD cases are not usually directly evaluated by psychiatrists or clinical psychologists affiliated with the study – instead, non-psychiatrist research assistants gather the primary data with expert review occurring based on written records or only in reliability sub-studies.

We propose to remediate all of these difficulties using existing samples from the Netherlands (1,000 cases with MDD and 1,000 matched controls). All subjects are participants in two large, longitudinal, population-based studies. The parent studies are closely coordinated, samples are from a known epidemiological sampling frame, controls are well-matched to cases, multiple biomarkers are available for cases and controls (diurnal cortisol variation, thyroid function, and autonomic nervous system assessment, fMRI, and lymphocyte microarrays). This is arguably the best and most comprehensive sample collection for MDD in the world.

3. Specific Aims of the Parent Studies

Cases and controls for Stage 1 GAIN genomewide association genotyping are from two longitudinal, population-based, and coordinated studies. Cases are from the *Netherlands Study of Depression and Anxiety (NESDA)* and controls from the *Netherlands Twin Registry (NTR)*; both studies are funded via Dutch and/or US NIH grants.

NESDA is a prospective cohort study that follows 2,850 persons aged 18-65 years at baseline, and after 1, 2, 4, and 8 years of follow-up. NESDA began in 2003 and recruitment will end in 12/2006. The NESDA specific aims are to examine:

- the long-term course of depressive and anxiety disorders;
- the genetic, biological, psychosocial, and somatic determinants of the long-term course of depressive and anxiety disorders; and
- the quality of care and treatment needs of these patients.

NTR is a longitudinal twin family study that began with comprehensive national birth records (1). The NTR enrolls twins, parents, siblings, and spouses for genetic-epidemiological studies. Seven large scale screenings from 1991-2006 yielded data on >19,000 subjects (8,808 twins, 2,825 siblings, 5,734 parents, and 1,889 spouses). The overarching aims of the NTR focus on the genetic and environmental determinants of:

- psychiatric disorders/mental health (including MDD) and personality;
- health-related behavior (exercise, nicotine, alcohol, drug use);
- physical health; and
- biomarkers for psychiatric and cardiovascular disease.

4. Disease/Trait (Major Depressive Disorder or MDD)

The defining features of MDD are marked and persistent dysphoria associated with physical and cognitive signs and symptoms (anhedonia, sleep disturbance, weight/appetite changes, motor agitation/retardation, anergia, excessive guilt or worthlessness, poor concentration or indecisiveness, and recurrent thoughts of death or suicide) (2). MDD is distinct from the normal sadness by its persistence (i.e., ≥ 2 weeks), the presence of these signs and symptoms, and its substantial associated impairment.

Combining available estimates, the lifetime prevalence of MDD is approximately 15% and is twofold higher in women (3). The course of MDD is typified by recurrence of illness: in a meta-analysis, 76% had ≥ 1 recurrence over a 10 year follow-up (4). MDD is associated with considerable morbidity (greater than chronic medical conditions such as diabetes and arthritis) (5-7), excess mortality from suicide and other causes (8-11), and substantial direct and indirect costs ($> \$43$ billion/year in the US) (12). A WHO study projected MDD to be the second leading cause of disability worldwide by 2020 (13).

Efforts to identify susceptibility loci for MDD have not been successful despite the presence of a solid rationale for the search. If an unbiased and well-powered Stage 1 genomewide association study were to identify risk or protective loci and if these were to be compellingly replicated, this new knowledge would have multiple critical ramifications. These data could lead to a revolution in our understanding of the etiology of MDD, suggest new etiological hypotheses (and perhaps novel drug targets), provide an essential means by which to stratify clinical samples, and alter the societal conceptualization of this highly significant complex trait.

All subjects in both NESDA and NTR have been assessed using the CIDI (14-17), an internationally accepted structured psychiatric diagnostic interview. The CIDI yields lifetime diagnoses of DSM-IV MDD (2) along with exclusionary conditions (e.g., history of stroke or moderate/severe drug or alcohol dependence). Age of onset and number of episodes are obtained (which may identify subgroups of MDD with greater heritability) (18) along with data on clinically important comorbid disorders. For those who screen positive for lifetime MDD in NESDA, psychiatrists or clinical psychologists affiliated with NESDA conduct semi-structured clinical evaluations in order to confirm that all inclusion and exclusion criteria are satisfied. In the NTR, longitudinal assessments are available and inclusion as a control requires the absence of MDD at multiple time points.

The following table summarizes the data available in NESDA and NTR. "Initial" indicates phenotypes that we commit to making available for analysis with the genomewide association genotypes from GAIN. "Future" indicates data that we commit to making available to NCBI when they become available in the future. This set of phenotypes is exceptionally comprehensive and is arguably superior to any other sample collection in the world. Note that the childhood trauma data may be an environmental risk factor that interacts with genetic susceptibility (19, 20). For almost all measures, assessments are highly congruent across studies (i.e., identical instrumentation or protocols). All assessments were selected after extensive literature reviews and expert consultation.

Some of the more expensive measures (MRI and microarrays) are available on subsets and we are seeking supplemental funding to increase the number of assessments.

Phenotype	Longitudinal	NESDA Cases	NTR Controls	NCBI Deposition
Demography – age, sex, ancestry, marital status, & education	Yes	Yes	Yes	Initial
CIDI – MDD information (episodes & age of onset)	Yes	Yes	n/a	Initial
Depression severity (Inventory of Depressive Symptoms)	Yes	Yes	n/a	Initial
Licit & illicit substance use	Yes	Yes	Yes	Initial
Personality (neuroticism & extraversion) (21, 22)	Yes	Yes	Yes	Initial
Anxiety severity	Yes	Yes	Yes	Initial
Prospective assessment of course of illness	Yes	Yes	n/a	Future
Trauma exposure in child hood & stressful life events	Yes	Yes	Yes	Future
Family history of MDD	Yes	Yes	Yes	Future
Thyroid function (TSH & free T ₃ , 99% of subjects)	Yes	Yes	No	Future
Cortisol profile (six time points, 75% of subjects)	Yes	Yes	Yes	Future
Heart Rate Variability (an index of autonomic nervous system tone via VU-AMS system, 95% of subjects)	Yes	Yes	Yes	Future
Lymphocyte microarrays (±LPS challenge, ~200 subjects)	Yes	Yes	Yes	Future
Brain MRI (with Tower of London, Eckman faces, & memory tests, ~200 subjects)	Yes	Yes	Yes	Future

5. Additional Phenotyping

NESDA and the NTR are longitudinal studies. We commit to making updated phenotypic data available to NCBI along with multiple potentially highly informative biomarkers.

6. Evidence for a Genetic Component

MDD is a complex trait and oligo- to polygenic inheritance is likely. Interactions at multiple levels (GxG, GxE) have been hypothesized. Early empirical data support GxE interactions with stressful/traumatic life events (19, 20).

Our meta-analysis of the genetic epidemiological literature on MDD (18) is included in the Appendix. Five *family studies* met inclusion criteria. The odds ratios for proband versus first-degree relative status) were homogeneous across the five studies (Mantel-Haenszel OR=2.84, 95% CI=2.31-3.49). No *adoption study* met inclusion criteria, but the results of two of the three reports were consistent with genetic influences on liability to MD. Five *twin studies* met inclusion criteria – familial aggregation was due to additive genetic effects - heritability of liability=37% (95% CI=31-42%), minimal shared environmental effects (0%, 95% CI=0-5%), and substantial individual-specific environmental effects/measurement error (63%, 95% CI=58-67%). MD is familial, and its familiarity mostly results from genetic influences.

Critically, the data also show that longitudinal assessment of MDD is essential. If MDD is measured longitudinally – as in both NESDA and NTR – then heritability increases substantially to ~60% (23-25).

The Figure below depicts all published genomewide linkage studies (26-33) of MDD or the related personality trait of neuroticism (21). There is relatively poor overlap for these studies consistent with the insufficiently low power of the individual studies (34).

7. Number of Samples

1,000 cases with MDD from NESDA plus 1,000 matched controls from the NTR will be available on 1 September 2006.

NOTE. Ascertainment in NESDA is on-going – there are ~1,000 cases with MDD now recruited but there is a slight lag in cleaning/incorporating data from recent subjects. Therefore, we have deposited with the NCBI data on N=783 cases and N=860 controls.

8. Project Design

All cases are from NESDA, a prospective cohort study following 2,850 persons with MDD or anxiety disorders aged 18-65 years at baseline, and after 1, 2, 4, and 8 years of follow-up. All controls are from NTR, a population-based twin-family study (total N>19,000 subjects). Using longitudinal data, controls are both unaffected and at low liability for MDD – an extremely strong case can be made for neuroticism as an endophenotype for MDD base on state correlations (35-39), prospective prediction (24, 40-43), and from multivariate twin genetic models (24, 36, 44, 45).

We propose to use a multistage design to minimize genotyping cost while preserving power and to maximize chances of true positive findings. Multistage designs are widely viewed as the optimal approach for genomewide association (46, 47).

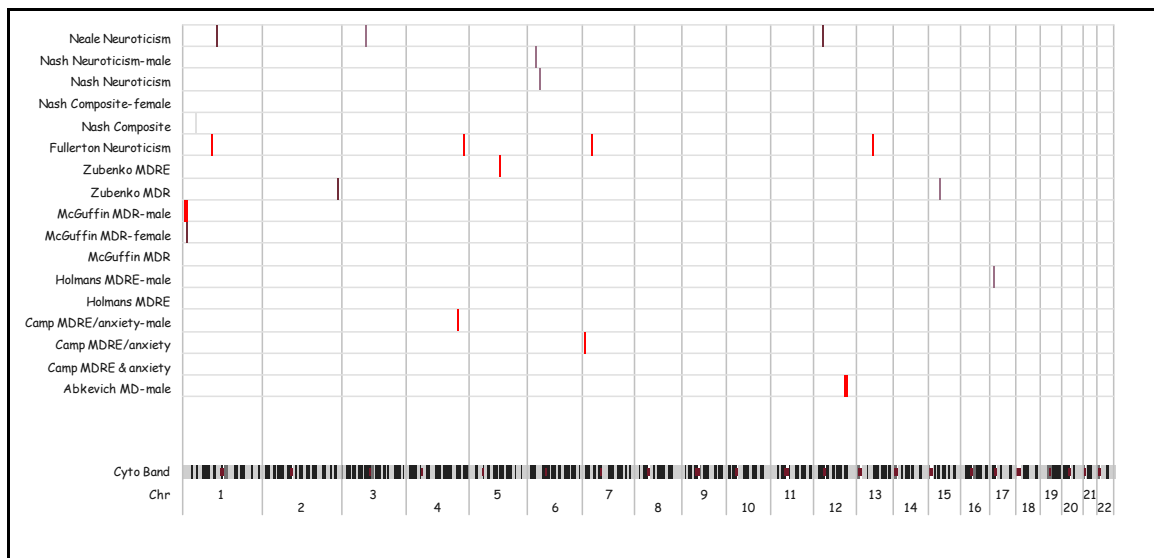


Figure 1. All published genomewide linkage studies of MDD or the endophenotype of neuroticism (24, 35-43). X-axis, human genome from chr1p to chr22q. Rows show the individual studies, and the highlighted regions show significant findings (LOD or equivalents ≥ 1.5). Figure by PFS based on hg16

In *Stage 1*, 1,000 NESDA cases and 1,000 matched NTR controls will be genotyped for a genomewide panel of SNP markers via GAIN. In *Stage 2*, we will attempt to genotype ~4,500 SNPs in two independent samples using Illumina GoldenGate assays.

Replication sample 1 are independent NTR cases and controls not included in Stage 1. Replication sample 2 are independent cases and controls from the Australian Twin Register. Each replication sample will contain 1,000 cases with MDD and 1,000 controls. All subjects are unrelated, and phenotypic classification is based on longitudinal assessments. (We have previously shown the comparability of the Dutch and Australian twin samples – based on 369 microsatellite markers, we calculated that the overall F_{st} between NTR- Australian samples was only 0.46% compared to 0.7% for European samples and 5% across multiple worldwide samples (48).

We have discussed replication with investigators on large studies of MDD – e.g., STAR*D (49), the NIMH Genetics Initiative collection, and Dr. Jonathan Flint (26, 50). Should we be selected for GAIN, we would convert these to formal replication protocols.

Unique features of the Stage 1 samples:

- Homogeneous subject ancestries, closely parallel parent studies
- Population-based sampling, cases and controls drawn from the same population
- Inherently longitudinal design
- Controls with no MDD and at low liability per longitudinal evaluations
- Multiple biological and exposures phenotypes available on cases and controls.
- MDD cases are evaluated by NESDA psychiatrists or clinical psychologists.

9. Power & Effect Size

We conducted two-stage genomewide association power calculations using CaTS (51) for joint analysis, four genetic models, and three allele frequencies. The minimum detectable genotypic relative risks with 80% power are shown in **Table 1**. The proposed study is well-powered to detect genotypic relative risks in the parameter space that is consistent with common disease-common variant models for complex traits. For psychiatric disorders, replicated candidate gene associations for schizophrenia (52) have relative risks in the 1.8-2.2 range. Therefore, we anticipate excellent power for this study. If MDD is caused by multiple rare variants, then power will be inadequate and genetic dissection will await the next generation of sequencing technologies.

Table 1: Minimum Detectable Genotypic Relative Risk (80% Power)

Allele frequency	Multiplicative	Additive	Dominant	Recessive
0.50	1.26	1.30	1.71	1.53
0.25	1.29	1.32	1.43	2.16
0.10	1.43	1.44	1.49	5.25

Assumptions: final $\alpha=0.05$; 2,000 cases and 2,000 controls (half genotyped in Stage 1); 375,000 SNPs genotyped in Stage 1 and 4,500 in Stage 2; SNP $\alpha=0.0000002$, based on prior work with HapMap2 data (53), we assume that the 375,000 SNPs actually genotyped are equivalent to 300,000 independent tests ($=375,000 \times 0.8$), yielding individual SNP $\alpha=1.7e-7 = 1-(1-0.05)^{1/300,000}$, and lifetime prevalence of MDD = 0.15 (3).

10. Study Population

NESDA Cases. NESDA is a longitudinal cohort study of 2,850 adults aged 18-65 years with MDD or anxiety disorders. In primary care, 850 subjects with MDD and/or an anxiety disorder (2) were recruited through a three-stage screening procedure – a mailed screener (K10) (16), a telephone CIDI interview (14), and a confirmatory interview with a NESDA psychiatrist or clinical psychologist. In secondary care, 850 respondents with MDD and/or an anxiety disorder (2) were recruited when starting treatment at a mental health care institution. All subjects provide written informed consent and the protocol has been reviewed and approved by multiple IRBs. The inclusion criterion for these 1,700 participants is a current DSM-IV diagnosis of MDD or anxiety disorder as confirmed by the CIDI (14). (An additional 1,150 subjects are included for other purposes.). A total of ~1,200 will have MDD. As of 6/1/2006, ~900 subjects with MDD will have been enrolled with DNA available. NESDA enrollment will be completed in 12/2006 and recruitment has been ~120 subjects/month, so we are highly confident of making the target of 1,000 cases with MDD (in fact, we may be able

to deliver 20-25% more subjects). Exclusion criteria: primary diagnoses of psychosis, bipolar disorder, obsessive compulsive disorder, severe drug/alcohol dependence, mood disorder due to a general medical condition, insufficient knowledge of the Dutch language, or ancestry other than Northern European. Sampling frame. (a) Primary care patients are recruited from 60 general practitioners in the vicinity of the three Universities involved, using a three-stage screening procedure. All GP patients aged 18-65 years who consulted the GP in the last 4 months are sent a short questionnaire (16). If GP patients score positive on this screener instrument (K10 score >19), they are telephoned for a brief interview (CIDI-short form for MDD and anxiety disorders). All GP patients who screen positive are invited for the NESDA baseline psychiatric interview. (b) The specialized mental health patients are recruited from outpatient regional facilities for mental health care. All newly registered patients with a primary diagnosis of MDD who fulfill inclusion and exclusion criteria, are referred to NESDA.

NTR Controls. The NTR enrolls twins, their parents, siblings and spouses for genetic-epidemiological studies of health and behavior. In six surveys (1991-2005) data were collected on physical and mental health; lifestyle; personality and mood; demographics and other variables. Data have been collected on >19,000 subjects (8,808 twins, 2,825 siblings, 5,734 parents and 1,889 spouses). Longitudinal phenotypes include MDD assessed using the CIDI (14) along with self-reported depressive symptoms (1).

Matching. Controls are group-matched to cases by sex and five year age band. Inclusion criteria: DNA available, no diagnosis of MDD on any interview occasion, at low liability for MDD via genetic factor scores/persistently low neuroticism (1), and unrelated to any other control. Exclusion criteria are identical to those for cases from NESDA. Sampling frame. The NTR begins with Dutch national records of multiple births and expands to include parents, siblings, and spouses.

For both samples, the epidemiological frame is known, and data will be available to estimate ascertainment bias. See Population Study Table in Section E. All subjects are of Northern European ancestry. Matching effectiveness: cases 72% female, mean age 41 years and controls 68% female and mean age of 39 years.

11. Data Management

Professional data managers maintain the databases for both NESDA and NTR. Both studies are on-going and are managed by Dr. Jan Smit (NESDA) and Dr. J.J. Hottenga (NTR). Data management software includes MySQL, SPSS, and SAS (54). Access is limited to the personnel of the parent studies. Database formats are general and there are no aspects that would need to be replicated in the NCBI version.

12. Data Analysis

Statistical Geneticist. We have been fortunate to attract UNC biostatistician Dr. Danyu Lin as the statistical geneticist for the genomewide association data. Dr. Lin is working on a body of methodological work on genomewide association and linkage studies (55-58). Dr. Lin will be assisted by Naomi Wray, PhD (statistical analysis) and the PI will assist with bioinformatics (the PI has been a professional SAS programmer for 30 years, and see TAMAL, <http://neoref.ils.unc.edu/tamal>, for an example) (59).

Analytic Plan. All statistical analyses will proceed according to a master data analysis plan what will be posted on the web prior to beginning data analysis (and following attendance at the GAIN data analysis workshop). In this way, it will be clear which analyses are pre-specified and which are *post-hoc*. The current consensus is that joint analysis is more powerful (51). Software includes SAS (54) and R (60). Our principal aim

is to investigate genetic associations relevant to the etiology of MDD via case-control comparisons, and the analytic plan is described below. We have also envisioned a thoughtful approach to the additional phenotypes relevant to delineating etiological subgroups, GxE interactions, and incorporating biomarker data. As we currently envision these as secondary analyses, we do not describe them here.

Initial Analyses. First, we will investigate ascertainment bias in cases and controls and conduct exploratory phenotypic analyses to confirm the adequacy of matching and establish the need for any covariates. Second, quality control – careful review of QC indices from the genotyping platform, HWE calculation (61), and no-call proportions by marker and by subject (62). Third, we will attempt to detect copy number polymorphisms in all subjects (63) (with the exact method contingent on the SNP genotyping platform). Fourth, we will investigate the presence and impact of population stratification and assignment of ancestry using genotype data via multiple empirical strategies (48, 64-69).

Stage 1. The central intention underlying this proposal is to compare genotype and haplotype frequencies in well-matched cases and controls with and without lifetime MDD. There are two key analyses – single SNP associations using logistic regression (dependent variable case/control status and independent variables of genotype (2 df test) plus any covariates) along with haplotype analyses with Dr. Lin's HAPSTAT (70).

Prioritization. We will prioritize genomic regions for genotyping in Stage 2. Inclusion will be based on (a) the strength of statistical signal from Stage 1, (b) the intersection of Stage 1 SNP findings with those of external genomewide association screens for MDD, and (c) candidate regions implicated by genomewide linkage or copy number screens from studies focusing on MDD. Selection will be explicitly operationalized.

Stage 2. Although we apply to GAIN for Stage 1 genotyping only, we have future plans. In Stage 2, we will genotype ~4,500 SNPs in two independent samples using Illumina GoldenGate assays. Replication sample 1 contains independent cases and controls from the NTR not included in Stage 1. Replication sample 2 are independent cases and controls from the Australian Twin Register. These two replication samples will each contain 1,000 cases with MDD and 1,000 controls. All subjects are unrelated, and phenotypic classification is based on longitudinal assessments. We have previously shown the comparability of the Dutch and Australian twin samples – based on 369 microsatellite markers, we calculated that the overall F_{st} between the NTR and the Australian Twin Register was only 0.46% compared to 0.7% for European samples and 5% across multiple worldwide samples (48).

We have engaged in tentative discussions with investigators on large studies of MDD – e.g., STAR*D (49), the NIMH Genetics Initiative collection, and Jonathan Flint in the UK (26, 50). Should we be selected for GAIN genotyping, we would convert these to formal replication protocols.

Empirical P-Values & Adjustment for Multiple Testing. Many SNPs will be in LD, and we need to construct tests that have the correct Type I error. A common approach for evaluating statistical significance in the presence of LD is the use of permutation tests (71). Dr. Lin has described a Monte Carlo-based algorithm for empirically evaluating the statistical significance of findings from a two-stage design (GAS2 software) (55). This approach has been shown to provide valid genomewide significance levels and was considerably more powerful than Bonferroni correction.

Why should these samples/data be selected for GAIN? We believe this to be arguably the best dataset in the world for genomewide association for MDD. Unique features:

- Homogeneous subject ancestries (N. European) and parallel parent studies
- Population-based sampling
- Inherently longitudinal design
- Cases and controls drawn from the same population
- Multiple informative phenotypes available on cases and controls.
- Controls with no MDD per longitudinal evaluations
- MDD cases are evaluated by NESDA psychiatrists or clinical psychologists.

13. Replication

Two replication samples are currently available (with banked DNA and phenotypes). See Section A.12 above for details.

14. Follow-Up Genetic & Functional Studies

Under a CDCV model, the key initial task is to ensure that any genotype-phenotype association replicates convincingly in multiple samples. For highly compelling replications, we would pursue a multi-pronged approach.

The extensive and coordinated biobanking of sera, plasma, cell lines, MRI scans, and urine in NESDA and NTR mean that many follow-up studies could begin immediately. First, we would conduct nested sub-studies of NESDA subjects with and without a risk polymorphism to understand its phenotypic correlates (at the level of gene expression, protein alterations, intermediate phenotypes, brain imaging, and macro-level subject data). All could be done to a high technical standard using existing data and could be accomplished very quickly. Second, there will undoubtedly be additional data that we would want to collect in a nested sub-samples. As NESDA is longitudinal, we will be in continuing contact with subjects and can readily invite individuals for additional phenotypic assessments. In particular, a biopsy of olfactory epithelium (a 10 minute ENT procedure) can be used to establish neuronal cell lines derived from CNS. Third, we would extend data collection to the first-degree relatives better to understand familial and genetic modes of transmission (this may be critical if copy number polymorphism studies suggest multiple rare variants). Fourth, UNC/Genetics is one of the top mouse genetics departments in the US and excellent transgenic and phenotyping core facilities exist and could readily create and evaluate the appropriate mouse model.

15. Source of DNA & Extraction Method

In both NESDA and NTR, all DNA samples were extracted from peripheral blood lymphocytes using PUREGENE kits (Gentra, Minneapolis, MN).

16. Availability of DNA Samples

DNA samples can be shipped by 1 September 2006.

17. Previous Genotyping

Subjects in NESDA have not yet been genotyped. NTR subjects have been genotyped on multiple occasions (including a genomewide linkage panel by Marshfield, in preparation).

References

1. Boomsma DI, Beem AL, van den Berg M, Dolan CV, Koopmans JR, Vink JM, et al. Netherlands twin family study of anxious depression (NETSAD). *Twin Res* 2000;3(4):323-34.
2. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Fourth Edition ed. Washington, DC: American Psychiatric Association; 1994.
3. Weissman MM, Bland R, Joyce PR, Newman S, Wells JE, Wittchen H-U. Sex differences in rates of depression: cross-national perspectives. *Journal of Affective Disorders* 1993;29:77-84.
4. Piccinelli M, Wilkinson G. Outcome of depression in psychiatric settings. *British Journal of Psychiatry* 1994;164:297-304.
5. Wells KB, Stewart A, Hays RD, Burnam MA, Rogers W, Daniels M, et al. The functioning and well-being of depressed patients: results from the Medical Outcomes Study. *Journal of the American Medical Association* 1989;262:914-919.
6. Broadhead WE, Blazer DG, George LK, Tse CK. Depression, disability days, and days lost from work in a prospective epidemiologic survey. *Journal of the American Medical Association* 1990;264:2524-2528.
7. Judd LL, Paulus MP, Wells KB, Rapaport MN. Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population. *American Journal of Psychiatry* 1996;153:1411-1417.
8. Tsuang MT, Woolson RF. Excess mortality in schizophrenia and affective disorders. *Archives of General Psychiatry* 1978;35:1181-1185.
9. Berglund M, Nilsson K. Mortality in severe depression: a prospective study including 103 suicides. *Acta Psychiatrica Scandinavica* 1987;76:372-380.
10. Black DW, Winokur G, Nasrallah A. Is death from natural causes still excessive in psychiatric patients? *Journal of Nervous and Mental Disease* 1987;175:674-680.
11. Zilber N, Schufman N, Lerner Y. Mortality among psychiatric patients - the groups at risk. *Acta Psychiatrica Scandinavica* 1989;79:248-256.
12. Greenberg PE, Stiglin LE, Finkelstein SN, Berndt ER. The economic burden of depression in 1990. *Journal of Clinical Psychiatry* 1993;54:405-418.
13. Murray CJL, Lopez AD. Evidence-based health policy: lessons from the Global Burden of Disease Study. *Science* 1996;274:740-743.
14. Robins LN, Wing J, Wittchen H-U, Helzer JE. The Composite International Diagnostic Interview: an epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry* 1988;45:1069-1077.
15. Kessler RC, Wittchen H-U, Abelson JM, McGonagle K, Schwarz N, Kendler KS, et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey (NCS). *International Journal of Methods in Psychiatric Research* 1997;7:33-55.

16. Kessler RC, Barker PR, Colpe LJ, Epstein JF, Gfroerer JC, Hiripi E, et al. Screening for serious mental illness in the general population. *Arch Gen Psychiatry* 2003;60(2):184-9.
17. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, et al. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). *Int J Methods Psychiatr Res* 2004;13(2):122-39.
18. Sullivan PF, Neale MC, Kendler KS. The genetic epidemiology of major depression: review and meta-analysis. *American Journal of Psychiatry* 2000;157:1552-1562.
19. Caspi A, Sugden K, Moffitt TE, Taylor A, Craig IW, Harrington H, et al. Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 2003;301(5631):386-9.
20. Kendler KS, Kuhn JW, Vittum J, Prescott CA, Riley B. The interaction of stressful life events and a serotonin transporter polymorphism in the prediction of episodes of major depression: a replication. *Arch Gen Psychiatry* 2005;62(5):529-35.
21. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Personality and Individual Differences* 1985;6:21-29.
22. Costa PT, McCrae RR. *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources; 1985.
23. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A longitudinal twin study of 1-year prevalence of major depression in women. *Archives of General Psychiatry* 1993;50:843-852.
24. Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ. A longitudinal twin study of personality and major depression in women. *Archives of General Psychiatry* 1993;50:853-862.
25. Kendler KS, Kessler RC, Neale MC, Heath AC, Eaves LJ. The prediction of major depression in women: toward an integrated etiologic model. *American Journal of Psychiatry* 1993;150:1139-1148.
26. Fullerton J, Cubin M, Tiwari H, Wang C, Bomhra A, Davidson S, et al. Linkage analysis of extremely discordant and concordant sibling pairs identifies quantitative-trait loci that influence variation in the human personality trait neuroticism. *American Journal of Human Genetics* 2003;72:879-90.
27. Abkevich V, Camp NJ, Hensel CH, Neff CD, Russell DL, Hughes DC, et al. Predisposition locus for major depression at chromosome 12q22-12q23.2. *Am J Hum Genet* 2003;73(6):1271-81.
28. Zubenko GS, Maher B, Hughes HB, 3rd, Zubenko WN, Stiffler JS, Kaplan BB, et al. Genome-wide linkage survey for genetic loci that influence the development of depressive disorders in families with recurrent, early-onset, major depression. *Am J Med Genet B Neuropsychiatr Genet* 2003;123(1):1-18.
29. Neale BM, Sullivan PF, Kendler KS. A genome scan of neuroticism in nicotine dependent smokers. *Am J Med Genet B Neuropsychiatr Genet* 2004;132B:65-69.

30. Nash MW, Huezo-Diaz P, Williamson RJ, Sterne A, Purcell S, Hoda F, et al. Genome-wide linkage analysis of a composite index of neuroticism and mood-related scales in extreme selected sibships. *Hum Mol Genet* 2004;13(19):2173-82.
31. McGuffin P, Knight J, Breen G, Brewster S, Boyd PR, Craddock N, et al. Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Hum Mol Genet* 2005;14(22):3337-45.
32. Holmans P, Zubenko GS, Crowe RR, DePaulo JR, Jr., Scheftner WA, Weissman MM, et al. Genomewide significant linkage to recurrent, early-onset major depressive disorder on chromosome 15q. *Am J Hum Genet* 2004;74(6):1154-67.
33. Camp NJ, Lowry MR, Richards RL, Plenk AM, Carter C, Hensel CH, et al. Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *Am J Med Genet B Neuropsychiatr Genet* 2005;135(1):85-93.
34. Risch N, Merikangas KR. The future of genetic studies of complex human diseases. *Science* 1996;273:1516-1517 Erratum: 1997, 275, 1329-1330.
35. McGuire RJ, Mowbray RM, Vallance RC. The MPI used with psychiatric inpatients. *British Journal of Psychology* 1963;54:157-166.
36. Wetzel RD, Cloninger CR, Hong B, Reich T. Personality as a subclinical expression of the affective disorders. *Comprehensive Psychiatry* 1980;21:197-205.
37. Hirschfeld RMA, Klerman GL. Personality attributes and affective disorders. *American Journal of Psychiatry* 1979;136:67-70.
38. Kendell RE, DiScipio WJ. EPI scores of patients with depressive illnesses. *British Journal of Psychiatry* 1968;114:767-770.
39. Bech P, Jørgensen B, Jeppesen K, Poulson DL, Vanggaard T. Personality in depression: concordance between clinical assessment and questionnaires. *Acta Psychiatrica Scandinavica* 1986;74:263-268.
40. Angst J, Clayton PJ. Premorbid personality of depressive, bipolar, and schizophrenic patients with special reference to suicidal issues. *Comprehensive Psychiatry* 1986;27:511-532.
41. Nystrom S, Lindegard B. Predisposition for mental syndromes. *Acta Psychiatrica Scandinavica* 1975;51:69-76.
42. Hirschfeld RMA, Klerman GL, Lavori PW, Keller MB, Griffith P, Coryell W. Premorbid personality assessments of first onset of major depression. *Archives of General Psychiatry* 1989;46:345-350.
43. Boyce P, Parker G, Barnett B, Cooney M, Smith F. Personality as a vulnerability factor to depression. *British Journal of Psychiatry* 1991;159:106-114.
44. Krieg JC, Lauer CJ, Hermie L, von Bardeleben U, Pollmacher T, Holsboer F. Psychometric, polysomnographic, and neuroendocrine measures in subjects at high risk for psychiatric disorders: preliminary results. *Neuropsychobiology* 1990;23:57-67.
45. Maier W, Lichtermann D, Mingos J, Heun R. Personality traits in subjects at risk for unipolar major depression: a family study perspective. *Journal of Affective Disorders* 1992;24:153-164.

46. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. *Am J Hum Genet* 2005;77(3):337-45.
47. van den Oord EJ, Sullivan PF. A framework for controlling false discovery rates and minimizing the amount of genotyping in the search for disease mutations. *Hum Hered* 2003;56(4):188-99.
48. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, et al. Genetic structure of human populations. *Science* 2002;298(5602):2381-5.
49. McMahon FJ, Buervenich S, Charney D, Lipsky R, Rush AJ, Wilson AF, et al. Variation in the Gene Encoding the Serotonin 2A Receptor Is Associated with Outcome of Antidepressant Treatment. *Am J Hum Genet* 2006;78(5):804-14.
50. Flint J. The genetic basis of neuroticism. *Neurosci Biobehav Rev* 2004;28(3):307-16.
51. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006.
52. Sullivan PF. The genetics of schizophrenia. *PLoS Medicine* 2005;2:614-618.
53. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature* 2005;437(7063):1299-320.
54. SAS Institute Inc. SAS/STAT® Software: Version 9. Cary, NC: SAS Institute, Inc.; 2004.
55. Lin DY. Evaluating statistical significance in two-stage genomewide association studies. *Am J Hum Genet* 2006;78(3):505-9.
56. Diao G, Lin DY. A powerful and robust method for mapping quantitative trait loci in general pedigrees. *Am J Hum Genet* 2005;77(1):97-111.
57. Lin DY. An efficient Monte Carlo approach to assessing statistical significance in genomic studies. *Bioinformatics* 2005;21(6):781-7.
58. Lin DY, Zou F. Assessing genomewide statistical significance in linkage studies. *Genet Epidemiol* 2004;27(3):202-14.
59. Hemminger BM, Saelim B, Sullivan PF. TAMAL: An integrated approach to choosing SNPs for genetic studies of human complex traits. *Bioinformatics* 2006;22:626-7.
60. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2005.
61. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of hardy-weinberg equilibrium. *Am J Hum Genet* 2005;76(5):887-3.
62. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;21:263-5.
63. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. *Science* 2004;305(5683):525-8.
64. Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999;55:997-1004.

65. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, Thomas MG, et al. Population genetic structure of variable drug response. *Nat Genet* 2001;29(3):265-9.
66. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet* 2003;72(6):1492-1504.
67. Bacanu S-A, Devlin B, Roeder K. The power of genomic control. *American Journal of Human Genetics* 2000;66:1933-1944.
68. Satten GA, Flanders WD, Yang Q. Accounting for unmeasured population substructure in case-control studies of genetic association using a novel latent-class model. *Am J Hum Genet* 2001;68(2):466-77.
69. Reich DE, Goldstein DB. Detecting association in a case-control study while correcting for population stratification. *Genet Epidemiol* 2001;20(1):4-16.
70. Lin DY, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. *Genet Epidemiol* 2005;29(4):299-312.
71. Good PI. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses*. New York: Springer-Verlag; 2004.